To: Unicode Technical Committee
From: Debbie Anderson, Script Encoding Initiative, UC Berkeley
Subject: Further information on Khitan Small Script clusters
Date: 31 August 2018

Background
In July, I received an email from Prof. Wu about Khitan Small Script cluster patterns that could impact the encoding model. (He had earlier shared this information at the London ISO JTC1 SC2/WG2 meeting, but images were not provided at the time.) KSS was on the CD ballot.

Below are Prof. Wu's examples and his description, followed two questions posed by Liang Hai over Twitter, with responses by Andrew West and V. Zaytsev, who were co-authors of the Khitan Small Script proposal.

Examples
Below are the examples provided by Prof. Wu, showing the clusters he wanted to call attention to:

Comments from Prof. Wu (translation by Liang Hai):
Images in the attachment are two different written forms of one *composite character* (translator's note: i.e., a cluster formed from simple characters), the latter one is a newly observed special structure, which can prove the former one are formed from 4 *original characters* (translator's note: i.e., simple character), instead of 3 original characters we previously thought of.

The following are the characters in the images:

18C5B   18CCF   18C90

18C5B   18CCC   18B42 18C90

Twitter discussion 9 August 2018

Liang Hai:

Are both clusters (provided by Prof Wu) well supported by the latest encoding model and character repertoire? (<…, 18CCF, …> vs <…, 18CCC, 18B42, …>?) Are there more homophonous pairs like this?

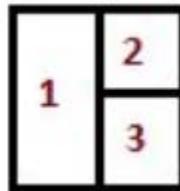V. Zaytsev:

Both appear to be standard clusters well supported by the proposed encoding model and character repertoire: <18C5B 18CCF 18C90> and <18C5B 18CCC 18B42 18C90>

Andrew West:

As I understand the explanation given by Prof. Wu in London, on the basis of the right example, he interprets 18CCF in the left example as equivalent to two phonetic elements (18CCC + 18B42). This phonetic interpretation does not affect the graphic structure of the cluster.

Andrew West:

Likewise, he interprets the word huduk 'good fortune' <18C5B 18CD1> as being constructed from 3 phonetic elements, with two phonetic elements combined into one. Firstly, it is only his theory; and secondly the phonetic interpretation does not affect the graphic representation.